BP

(54) **CHARACTERISING POLYPEPTIDES**

CHARAKTERISIERUNG VON POLYPEPTIDEN

PROCEDE PERMETTANT DE CARACTERISER DES POLYPEPTIDES

(72) Inventors:
     • SCHMIDT, Günter
       Cambs PE17 2BQ (GB)
     • THOMPSON, Andrew, Hugin
       Ayr KA7 4RH (GB)

(74) Representative: Daniels, Jeffrey Nicholas et al
     Page White & Farrer
     54 Doughty Street
     London WC1N 2LS (GB)

(56) References cited:
     EP-A- 0 306 610          EP-A- 0 594 164
     WO-A-90/05192            GB-A- 2 269 177

     • SEIELSTAD, DONALD A. ET AL: "Analysis of the
       Structural Core of the Human Estrogen Receptor
       Ligand Binding Domain by Selective
       Proteolysis/Mass Spectrometric Analysis"
       BIOCHEMISTRY (1995), 34(39), 12605-15
       CODEN: BICHAW;ISSN: 0006-2960,
       XP002063749

     • DEBRABANT, ALAIN ET AL: "Intramolecular
       mapping of Plasmodium falciparum P126
       proteolytic fragments by N - terminal amino acid
       sequencing" MOL. BIOCHEM. PARASITOL.
       (1992), 53(1-2), 89-95 CODEN: MBIPDP;ISSN:
       0166-6851, XP002063750

     • MAREYA, SHADRECK M. ET AL: "Mapping the
       structural domains of E. coli carbamoyl
       phosphate synthetase using limited
       proteolysis" BIOORG. MED. CHEM. (1995), 3(5),
       525-32 CODEN: BMECEP;ISSN: 0968-0896,
       XP002063751

     • MUIR T.W. ET AL: "Probing the chemical basis
       of binding activity in an SH3 domain by protein
       signature analysis" BIOLOGY, vol. 3, no. 10,
       October 1996, LONDON, pages 817-825,
       XP002063752

     • SANGER: 'Biochemistry', vol. 3, 1988 pages 50 -
       57

     • ALBERTS ET AL: 'Molecular Biology of the cell',
       vol. 3, page 173

Remarks:
     The file contains technical information submitted
     after the application was filed and not included in this
     specification

EP 1 027 454 B1

BEST AVAILABLE COPY

## Description

## Field of the Invention

[0001] The present invention relates to a method for characterising polypeptides and to methods for identifying and assaying such polypeptides.

## Background to the Invention

[0002] The characterisation and identification of polypeptides from complex mixtures thereof, such as protein samples found in biological systems, is a well-known problem in biochemistry. Traditional methods involve a variety of liquid phase fractionation and chromatography steps followed by characterisation, for example by two dimensional gel electrophoresis. Such methods are prone to artefacts and are inherently slow. Moreover, automation of such methods is extremely difficult.

[0003] Patent Application PCT/GB97/02403, filed on 5th September 1997, describes a method for profiling a cDNA population in order to generate a 'signature' for every cDNA in the population. It is assumed in that method that a short sequence of about 8 bp that is determined with respect to a fixed reference point is sufficient to identify almost all genes. This system relies on immobilising the cDNA population at the 3' terminus and cleaving it with a restriction endonuclease. This leaves a population of 3' restriction fragments. The patent describes a technique that allows one to determine a signature of roughly 8 to 10 base pairs at a specified number of bases from the restriction site which is a sufficient signature to identify nearly all genes.

[0004] Techniques for profiling proteins, that is to say cataloguing the identities and quantities of proteins in a tissue, are less well developed in terms of automation or high throughput. The classical method of profiling a population of proteins is by two-dimensional electrophoresis. In this method a protein sample extracted from a biological sample is separated on a narrow gel strip. This first separation usually separates proteins on the basis of their iso-electric point. The entire gel strip is then laid against one edge of a rectangular gel. The separated proteins in the strip are then electrophoretically separated in the second gel on the basis of their size. This technology is slow and very difficult to automate. It is also relatively insensitive in its simplest incarnations. A number of improvements have been made to increase resolution of proteins by 2-D gel electrophoresis and to improve the sensitivity of the system. One method to improve the sensitivity of 2-D gel electrophoresis and its resolution is to analyse the protein in specific spots on the gel by mass spectrometry. One such method is in-gel tryptic digestion followed by analysis of the tryptic fragments by mass spectrometry to generate a peptide mass fingerprint. If sequence information is required, tandem mass spectrometry analysis can be performed.

[0005] More recently attempts have been made to exploit mass spectrometry to analyse whole proteins that have been fractionated by liquid chromatography or capillary electrophoresis. In-line systems exploiting capillary electrophoresis mass spectrometry have been tested. The analysis of whole proteins by mass spectrometry, however, suffers from a number of difficulties. The first difficulty is the analysis of the complex mass spectra resulting from multiple ionisation states accessible by individual proteins. The second major disadvantage is that the mass resolution of mass spectrometers is at present quite poor for high molecular weight species i.e. for ions that are greater than about 4 kilodaltons in mass so resolving proteins that are close in mass is difficult. A third disadvantage is that further analysis of whole proteins by tandem mass spectrometry is difficult as the fragmentation patterns for whole proteins are extremely complex.

[0006] Biochemistry (1995) 34, 12605-12615 discloses analysis of the structural core of the human oestrogen receptor ligand binding domain by cleavage by proteases, and subsequent analysis of the protein fragments by SDS-PAGE, Edman N-terminal sequncing and mass spectrometry (HPLC-coupled electrospray ionisation mass spectrometry).

## Summary of the Invention

[0007] The present invention provides a method for characterising polypeptides, which comprises:

> (a) treating a sample comprising a population of a plurality of polypeptides with a cleavage agent which is known to recognise in polypeptide chains a specific amino acid residue or sequence and to cleave at a cleavage site, whereby the population is cleaved to generate peptide fragments;

> (b) isolating a first population of peptide fragments which comprises only terminal peptide fragments bearing as a reference terminus the N-terminus or the C-terminus of the polypeptide from a second population of peptide fragments which do not comprise the reference fragment each peptide fragment of the first population bearing at the other end the cleavage site proximal to the reference terminus, said isolation step comprising immobilisation of the first population of peptide fragments or immobilisation of the second population of peptide fragments onto a solid support, and

> (c) determining by mass spectrometry a signature sequence of at least some of the isolated fragments, which signature sequence is the sequence of a predetermined number of amino acid residues running from the cleavage site;

wherein a signature sequence characterises each polypeptide.

[0008] The invention therefore describes a system analogous to that of PCT/GB97/02403, but for use with proteins. Since there are 20 monomers that make up a protein there are a great many more possible variants at a particular site in a sequence and so the length of signature required from a protein sequence is much shorter than that required from a cDNA sequence to identify it uniquely.

[0009] This invention can use liquid phase separation techniques and mass spectrometry to resolve proteins and protein fragments to facilitate automation and avoid the artefacts and inherent slowness and lack of automation in gel based techniques such as 2-D gel electrophoresis.

[0010] The reference terminus may be attached to a solid phase support to immobilise the population of polypeptides or peptide fragments thereof. Preferably, the population of polypeptides is immobilised before treatment with the cleavage agent. In this way, the peptide fragments produced on treatment with the cleavage agent remain immobilised and can be readily isolated by washing away unwanted material present in the liquid phase. The solid phase support may comprise suitable beads or other such supports well known in this art. Such supports or substrates may be chosen to bind selectively to either the N-terminus or the C-terminus and this is discussed in further detail below.

[0011] In one embodiment, the reference terminus is attached to the solid phase support by: (i) treating the polypeptides with a blocking agent to block all exposed reference groups, which comprise either carboxyl groups or primary amine groups; (ii)cleaving the reference terminal amino acids to expose unblocked reference termini; and (iii) treating the unblocked reference termini with an immobilisation agent capable of coupling to the solid phase support; wherein step (b) comprises binding the treated refrence termini to the solid phase support and removing unbound peptide fragments. In an alternative embodiment, the method further comprises (i) preparing the sample step (a) by pre-treating the polypeptides with a blocking agent to block all exposed reference groups, which comprise either carboxyl groups or primary amine groups, so that subsequent treatment of the sample with the cleavage agent generates peptide fragments bearing unblocked reference termini; (ii) biotinylating the unblocked reference termini; and (iii) binding the peptide fragments containing the unblocked reference termini to a solid phase support; wherein step (b) comprises eluting unbound peptide fragments therefrom. Preferably, the immobilisation agent comprises a biotinylation agent.

[0012] The cleavage agent must recognise a specific amino acid residue or sequence of amino acids reliably. The cleavage site may be at the specific amino acid residue or sequence or at a known displacement therefrom. The cleavage agent may be a chemical cleavage agent such as cyanogen bromide. Preferably, the cleavage agent is a peptidase, such as a serine protease, prefer-

ably trypsin.

As discussed in further detail below, depending on the number of proteins or polypeptides in a given sample, it may be advantageous to Sort the polypeptides into manageable sub-populations. Sorting can be effected before treatment of the sample with the cleavage agent or after cleavage. As discussed in further detail below, the sample of step (a) may comprise a sub-cellular fraction. In this way, the method further comprises a step of sub-cellular fractionation before step (a). The sample of step (a) may be prepared by liquid chromatography of either a crude fraction or a sub-cellular fraction. A preferred method of determining the signature sequence is by mass spectrometry and this may be preceded by a high pressure liquid chromatography step to resolve the peptide fragments. Alternatively, the peptide fragments may be subjected to ion exchange chromatography before step (c) followed by sequencing by mass spectrometry.

[0013] In accordance with the method of the present invention, the predetermined number of amino acid residues required to constitute the signature sequence will vary according to the size of the polypeptide population. Preferably, the predetermined number of amino acid residues is from 3 to 30, more preferably 3 to 6.

[0014] The present invention further provides a method for identifying polypeptides in a test sample. The method comprises characterising the polypeptides as described above and comparing the signature sequences and relative positions of the cleavage site obtained thereby with the signature sequences and relative positions of the cleavage site of reference polypeptides in order to identify the or each polypeptide in the test sample. This method can be used to identify a single unknown polypeptide or a population of unknown polypeptides by comparing their characteristics (i.e. their signature sequences and relative positions of cleavage site) with those of previously identified polypeptides. It is envisaged that the database of such characteristic can readily be compiled.

[0015] In a further aspect, the present invention provides a method for assaying for one or more specific polypeptides in a test sample. The method comprises performing a method as described above, wherein the cleavage agent and relative position of the cleavage site is predetermined and the signature sequence is determined in step (c) by assaying for a predetermined sequence of amino acid residues running from the cleavage site. Preferably, the cleavage site and signature sequence are predetermined by selecting corresponding sequences from one or more known target polypeptides, such as those available from the database.

**Brief Description of the Drawings**

[0016] The invention will now be described in further detail, by way of example only, with reference to the accompanying drawings, in which:

FIGURE 1 shows a reaction scheme according to one embodiment of the invention;

FIGURE 2 shows a reaction scheme according to another embodiment of the invention;

FIGURE 3 shows a reaction scheme according to a simple embodiment of the invention; and

FIGURE 4 shows a reaction scheme according to a variation of the embodiment shown in Figure 1.

## Brief Description of the Invention

### Protein Signatures:

[0017] The essence of this system is that one can immobilise a population of proteins onto a solid phase substrate at one terminus of the molecule. Proteins are directional so a particular terminus can be chosen in a manner dependent on the chemistry of the immobilisation agent, for example the Edman reagent (phenyl isothiocyanate) can be used selectively to remove amino acids from the N-terminus of a protein; however, if phenyl isocyanate is used the N-terminus is simply capped. A derivative of this molecule that could be coupled to a cleavable linker on a solid-phase substrate would allow a protein to be immobilised at its N-terminus and subsequently removed by cleavage of the linker. During peptide synthesis, the C-terminus is usually immobilised as a benzyl ester, through the use of a chloromethyl group. Such chemistry may be adapted to immobilise proteins at this terminus, if desired.

[0018] A population of immobilised proteins is then treated with a sequence specific peptidase such as trypsin to leave a population of N-terminal cleavage fragments. Such fragments can be considered to be analogous to an expressed sequence tag for a protein. One can then sequence the resultant peptide signatures by mass spectrometry. Terminal fragments are most meaningful, in that the position of all resultant peptide in the protein is known and the termini are usually accessible at the surface of most proteins.

## Sorting Proteins:

[0019] Obviously a population of proteins extracted from a cell is going to be a significant number of distinct species. If, as it is thought there are roughly 15000 genes expressed in the average human cell, one can expect as many proteins. Clearly one cannot sequence all of these by mass spectrometry in a single step, with present technology. For this reason a protein population of such size needs to be sorted into manageable subsets.

[0020] A generalised system for profiling proteins must attempt to resolve a protein population into reasonably discrete subsets of relatively uniform size. This is most readily achieved by separation on the basis of global properties of proteins, that vary over a broad and continuous range, such as size and surface charge,

which are the properties used most effectively in 2-D gel electrophoresis. Such separations can be achieved as rapidly or more so using liquid chromatographic techniques. In fact, by following one liquid chromatography separation by another, one can resolve proteins in as many dimensions as one requires, since there is a great deal more flexibility in liquid chromatography separation systems, although one would ideally avoid too many separation steps to prevent sample loss.

[0021] Sorting can be effected during extraction, after extraction of proteins from their source tissue or after cleavage of immobilised peptides.

### Sorting during cell fractionation:

[0022] Proteins are intrinsically sorted in vivo, in terms of their compartmentalisation within a cell. Various techniques are available that allow one to sort proteins on the basis of their cellular compartments. Fractionation protocols involve various cell lysis techniques such as sonication, detergents or mechanical cell lysis that can be coupled to a variety of fractionation techniques, mainly centrifugation. Separation into membrane proteins, cytosolic proteins and the major membrane bound sub-cellular compartments, such as the nucleus and mitochondria, is standard practice. Thus one can effectively ignore certain classes of protein if one chooses, e.g. mitochondrial proteins are likely to be uninteresting in a lot of cases. Membrane, cytosolic and nuclear compartments will be of particular interest on the whole.

### Sorting after extraction:

[0023] Since proteins are highly heterogenous molecules numerous techniques for separation of proteins are available on the basis of size, hydrophobicity, surface charge and various combinations of the above using liquid chromatography in its various incarnations. Separation is effected by an assortment of solid phase matrices derivitised with various functionalities that adhere to and hence slow down the flow of proteins through the column on the basis of the properties above. Molecules are normally loaded into such columns in conditions favouring adhesion to the solid phase matrix and selectively washed off in steadily increasing quantities of a second buffer favouring elution. In this way the proteins with the weakest interactions with a given matrix elute first.

[0024] Various formats for liquid chromatography exist but for greatest speed of throughput and for the most discrete separations High Pressure Liquid Chromatography (HPLC) formats are favoured. In this format the matrix is designed to be highly incompressible and when derivitised allow chromatographic separation to be performed at extremely high pressures which favours rapid and discrete separation.

## Sorting of cleaved peptides:

[0025] Liquid chromatography mass spectrometry (LCMS) is a well developed field. HPLC systems directly coupled to electrospray mass spectrometers are in widespread use. HPLC is a fast and effective way of resolving peptides after they have been cleaved from their immobilised state.

[0026] Alternatively sorting peptides by ion exchange chromatography might be advantageous, in that short peptides could be separated in an almost sequence dependent manner: the amino acids that are ionisable have known pKa values and hence elution of peptides from such a column at a specific pH, would be indicative of the presence of particular amino acids in that sequence. For example, aspartate residues have a pKa of 3.9 and glutamate residues 4.3. Elution of a peptide at pH 4.3 would be indicative of the presence of glutamate in the peptide. These effects are sometimes masked in large proteins but should be distinct in short peptides, hence would be extremely useful as sorting features.

[0027] Combination of the above techniques will allow various sorting protocols to be developed that will allow great control over the form of the protein profile generated. In this way, identification of most proteins expressed in a cell should be achievable.

## Sequencing of peptides by mass spectrometry:

[0028] Peptides can be readily sequenced directly by tandem mass spectrometry. In general, peptide mixtures are injected into the mass spectrometer by electrospray, which leaves them in the vapour phase. The first mass spectrometer acts as a filter selecting molecules to enter the second mass spectrometer on the basis of their mass charge ratio, such that essentially only a single species enters the second mass spectrometer at a time. On leaving the first mass spectrometer, the selected peptide passes through a collision chamber, which results in fragmentation of the peptide. Since fragmentation occurs mostly at the peptide bond, the pattern of fragments corresponds to a series of subspecies of peptides and amino acids that compose the original peptide. The distinct pattern of masses of single amino acids, 2-mers, 3-mers, etc. generated in the fragmentation of the peptide is sufficient to identify its sequence.

[0029] The end result is then that a population of proteins can be arbitrarily sorted into populations of peptides of convenient size to be fed into an electrospray tandem mass spectrometer for direct sequencing. Completion of such an analysis for an entire cell's proteins would give a profile of what proteins are present and in what relative quantities. Absolute quantitation could be achieved by 'spiking' a protein population with known quantities of particular proteins, known to be absent, e. g. plant proteins in animal samples or visa versa against which to calibrate results.

## Protein Signatures:

[0030] This invention provides a method of capturing a population of proteins onto a solid phase substrate by one terminus of each protein in the population. This invention also provides a method of cleaving proteins that have been derivatised at one terminus with an agent that can be used to immobilise that terminus on a solid phase substrate. This allows a single peptide for each protein in a population to be captured onto a solid phase substrate thus peptides from the chosen terminus can be separated from other peptides generated by the cleavage step and can be isolated. This invention also provides a method to allow all the peptides generated in a cleavage step that are not from the reference terminus to be captured leaving a single terminal peptide per protein free in solution for analysis.

[0031] A population of peptides generated according to the methods of this invention can be analysed in a number of ways preferably by mass spectrometry.

[0032] Two forms of analysis are preferred. The first is to determine peptide mass fingerprints for the population of signature peptides generated. In this method the mass of each peptide, preferably the accurate mass, is determined. A significant proportion of signature peptides should be uniquely identified by this form of analysis. Any mass peaks that are unknown can be further characterised by the second form of preferred analysis. Ions of a specific mass can be selected for collision induced dissociation in a tandem mass spectrometer. This technique can be used to determine sequence information for a peptide.

## Capturing Peptides:

[0033] This invention provides methods that exploit derivitisation of proteins with various agents, including existing peptide sequencing reagents, to isolate a single 'signature' peptide from each member of a population of proteins. This invention may be practised in two formats. The methods of this invention allow a reference terminus to be selected from the proteins in a population. In the first format this reference terminus may be derivatised with an immobilisation agent. If the proteins derivatised in this manner are treated with a sequence specific cleavage agent to generate peptides, the peptides from the reference termini of the proteins in a mixture can be specifically captured leaving the remaining peptides free in solution. This first format is discussed in the following section headed "Format 1" In the second format a single peptide sample per protein is generated by capturing the peptide fragments that are not from the chosen reference terminus, thus leaving the signature peptides free in solution.

## Format 1:

[0034] In the simplest embodiment of this invention as

shown schematically in Figure 3, a population of proteins is reacted with a modified sequencing agent specific for one terminus of each-protein in the population. The modified sequencing agent carries an immobilisation agent in order that proteins derivatised with the sequencing agent may be captured onto a solid phase support. The captured proteins may then be cleaved with a sequence specific cleavage agent. This cleavage step will generate a series of peptide fragments in solution and will leave a single peptide protein captured on the solid phase support. The peptides free in solution are then washed away. The immobilised peptides can then be released from the solid phase support by completing the sequencing reaction for the coupled terminal amino acid. The Edman reagent (phenyl isothiocyanate) could be modified to carry an immobilisation agent, the phenyl ring could be substituted with a group linked to an appropriate immobilisation effector such as biotin. A population of proteins derivatised with this reagent could be cleaved with trypsin. The derivatised terminal peptides could then be immobilised on an avidinated solid phase support allowing underivatised peptides to be washed away. The peptides could then be released from the solid phase support by disrupting the avidin-biotin reaction. This will leave N-terminal peptides free in solution. These peptides can then be analysed by mass spectrometry. It may be desirable to fractionate the peptides prior to mass spectrometry but this fractionation step is optional. Alternatively a modified C-terminal sequencing agent might be used to capture proteins by the C-terminus. The C-terminus is generally not post-translationally modified and so may be the preferred terminus to capture a population of proteins. Further embodiments of this invention are discussed below.

### C-terminal sequencing agents:

[0035] Unmodified C-terminal sequencing agents can be used to generate a signature peptide. A further embodiment of the present invention is as follows and is described schematically in Figure 1. In the first step a protein population extracted from a tissue is loosely immobilised onto a membrane, such as a PVDF membrane. The solvents used to extract proteins from a tissue sample are generally very harsh, usually containing agents such as urea, thiourea and detergents, since proteins have widely varying solubilities. Immobilising extracted proteins onto a membrane allows them to be washed with other solvents prior to modification. The protein population, thus captured, is then derivatised with a coupling agent, such as diphenyl phosphoroisothiocyanatidate from Hewlett-Packard, (Miller et al., Techniques in Protein Chemistry VI 219 - 227) in a method that is essentially the same as that which one would use for a normal sequencing reaction for a single protein giving peptidylacylisothiocyanates for all proteins. The coupling reagent also reacts with other free carboxyl groups also giving acylisothiocyanate derivatives. The

coupling agent may, however, react incompletely with some carboxylic acid side chains. It may, therefore, be desirable to perform additional derivitisation steps using more reactive reagents to ensure that all free carboxyl groups are derivatised. This variation is shown in Figure 4. The derivatised protein population is then treated with pyridine to effecting ring closure of the terminal acylisothiocyanate derivative. One can then cleave the C-terminal residue by addition of a cleavage agent such as trimethylsilanolate, from Hewlett-Packard, which cleaves the terminal amino acid from each protein releasing the thiohydantoin-amino acid derivative of the terminal amino acid. This exposes a free carboxyl at the penultimate residue of each protein. This can be specifically derivatised with biotin using 5 - (biotimamido) pentylamine since all other carboxyl groups are derivatised. In this way all the proteins in a population can be derivatised at the C-terminal with biotin. The biotinylated population, still on the PVDF membrane is then treated with an appropriate sequence specific cleavage agent. Trypsin is generally used for mass spectrometry applications as this generally leaves the N-terminal side of the cleavage site protonated which is desirable. Trypsin specifically cleaves adjacent to basic residues. If an enzyme is used the immobilised peptides would have to be washed with some form of physiological buffer to allow trypsin to function. This will leave a population of cleaved peptides, some of which are biotinylated which can be desorbed from the PVDF membrane into solution. The biotinylated peptides can be captured using a solid phase matrix derivatised with monomeric avidin. Non-immobilised peptides can then be washed away, leaving an immobilised population of C-terminal peptides which comprise the tag used to identify proteins in a population. After washing away free peptides, the immobilised tags can be released from the solid phase support by addition of acid which disrupts the biotin/avidin interaction - monomeric avidin is best for this purpose. In an alternative embodiment the biotinylated peptides can be captured on an avidinated support prior to sequence specific cleavage.

### N-terminal sequencing agents:

[0036] N-termini of a large proportion of cellular proteins are blocked. For the purposes of profiling those proteins whose N-termini are not blocked one can use the corresponding N-terminal sequencing agents to derivatise amino groups including the terminal amino group. The terminal amino acid can be cleaved and the newly exposed amine at the penultimate amino acid can be derivatised with an immobilisation agent. The biotinylated proteins can then be cleaved and the terminal signature peptides can be captured and analysed. This would however be limited to those N-termini that are not already blocked.

## Format 2:

[0037] This method is shown schematically in Figure 2. In this method a reagent that derivatises carboxyl residues is used to cap all carboxyl residues including the C-terminal carboxyl group in the protein population of interest. The protein population is then cleaved with trypsin or another sequence specific cleavage reagent that cleaves at the peptide bond to generate an amino and carboxyl group on the C-terminal and N-terminal fragments respectively. At this stage all peptides except the terminal peptides, which are capped will have a free carboxyl. These free carboxyls can be derivatised with 5 - (biotimamido)pentylamine or some other immobilisation agent. If biotin is used then one can capture all the biotinylated non C-terminal peptides onto a solid phase matrix derivatised with avidin. An avidinated affinity column in-line with a mass spectrometer would allow C-terminal peptides to be selectively eluted directly into the mass spectrometer for analysis.

[0038] This technique is equally applicable to generating peptide tags from the N-terminus of a population of proteins. Reagents which derivatise amine groups can be used to selectively cap all amine groups on a protein including the N-terminal amine group. Cleavage will expose amines in non-terminal peptides which can be derivatised with biotin allowing selective capture of non N-terminal peptides. This is important since many proteins are modified at the N-terminus and the N-terminal amine is often inaccessible to reagents. Thus selectively capturing non N-terminal peptides is a means of generating a signature at the N-terminus.

[0039] The reagents to derivatise amines and carboxyls are also simpler than those necessary for the coupling agents used in sequencing reactions.

## Immobilisation agents:

[0040] It is possible to capture derivatised peptides with a variety of chemical agents. In the discussion of the methods of this invention biotin has been chosen as an exemplary immobilisation agent due to its highly specific interactions with avidin. Other immobilisation agents besides biotin are compatible with the methods of this invention. The following are examples and the invention is not limited to these.

[0041] A linker to hexahistidine would allow peptide tags to be captured onto a coordinated metal ion derivatised column. Various antibody antigen interactions could be used as well where an antibody or antigen is tagged onto the penultimate amino acid rather than biotin.

## Antibodies against derivatives:

[0042] The most common N-terminal modification is acetylation. It should be possible to raise an antibody against N-terminally acetylated peptides to permit these

to be captured using an affinity column derivatised with such an antibody. In order to capture substantially all proteins one can derivatise the remaining proteins in a sample, that are not already acetylated, with an acetylation agent. The derivatised proteins can then be cleaved with chymotrypsin or another sequence specific agent (trypsin does not cleave acetylated cleavage sites of proteins). An anti-N-terminal acetylation antibody immobilised on an appropriate matrix could be used to generate an affinity column. Such a column could be used to capture peptide signatures with acetylated N-termini after their source proteins have been cleaved.

[0043] To capture C-terminal peptides one could raise an antibody against thiohydantoin derivatives of peptides which could be used to selectively capture a peptide from a protein that had been derivatised with a coupling agent for sequencing prior to cleavage with trypsin or another sequence specific cleavage agent.

## Derivitisation of proteins:

[0044] The methods of this invention include derivitisation steps which are required to ensure that the reference terminus of each protein in a population is specifically derivatised with an immobilisation agent in the first format or, in the second format, to ensure that the reference terminus is specifically blocked from reaction with an immobilisation agent. Additional derivitisation steps may also be performed. These may be desirable if fractionation of signature peptides is to be performed prior to mass spectrometry analysis. There are two important factors that should be considered with regard to any fractionation steps. These factors are the resolution of the fractionation step and the consequent sample loss imposed by the fractionation.

[0045] Certain chromatographic techniques are 'sticky' when used for the separation of peptides, that is to say a proportion of the sample is retained on the separation matrix. It is possible to reduce sample loss of this kind by derivitising the groups that are involved in adhesion to the separation matrix. That is to say, if one is using an ion exchange chromatography separation one can derivatise ionic and polar side chains with reagents that increase their hydrophobicity thus reducing affinity to the matrix. This will, however, reduce the resolution of the separation.

[0046] It is desirable to ensure that only one mass peak per peptide appears in the mass spectrum generated by analysis of a population of signature peptides. It may, therefore, be desirable to derivatise polar and ionic side chains of signature peptides in order to reduce the number of ionisation states accessible to those peptides. This step should help promote the formation of a single ion species per signature peptide.

[0047] It may also be desirable to add a group to each signature peptide to increase the sensitivity of the mass spectrometry analysis. A particularly good 'sensitising' group to add to a peptide would be a tertiary ammonium

ion which is a positively charged entity with excellent detection properties.

## Pre-Sorting Steps:

[0048] This technology can be used to profile peptide populations generated in numerous ways. Various fractionation techniques exist to sub-sort proteins on the basis of certain features. Of particular interest is the analysis of signalling pathways. Phosphorylation of proteins by kinases is a feature of many signalling pathways. Proteins that can be phosphorylated by a kinase often have a short phosphorylation motif that a kinase recognises. Antibodies exist that bind to such motifs, some binding phosphorylated forms while others bind the non-phosphorylated state. Antibody affinity columns or immuno-precipitation of kinase target sub-populations followed by profiling would be of great interest in identifying these proteins and in monitoring their metabolism simultaneously in time resolved studies of live model systems.

[0049] Many proteins exists as complexes and analysis of such complexes is often tricky. A cloned protein that is a putative member of a complex allows one to generate an affinity column with that protein to trap other proteins that bind to it. This profiling technology is eminently suited to analysis of such captured protein complexes.

Kits including antibody affinity columns to analyse signal transduction or membrane location by capturing proteins with the appropriate post-translational modifications are envisaged either as a pre-sorting step or as a capture step after cleavage of a protein population with a sequence specific cleavage agent.

## Chromatographic techniques:

[0050] Having generated peptide tags from a population of proteins it is then desirable to analyse the resultant tags. Chromatography is an optional step in the analysis of a population of peptide signatures prior to mass spectrometry but may be quite desirable depending on the configuration of the mass spectrometer used.

[0051] Two important features are required of any chromatographic stage in a protein profiling method, high resolution and minimal sample loss. Resolution generates information and also reduces the complexity of the peptide tag population entering the mass spectrometer. The second feature is that there is minimal loss of sample in the chromatographic separation, that would reduce the sensitivity of the technique to low frequency peptides in the population under analysis.

### Derivitisation of proteins:

[0052] Certain chromatographic techniques are 'sticky' when used for the separation of peptides, that is to say a proportion of the sample is retained on the sep-

aration matrix. To reduce sample loss of this kind is possible by derivitising the groups that are involved in adhesion to the separation matrix. That is to say, if one is using an ion exchange chromatography separation one can derivitise ionic and polar side chains with reagents that increase their hydrophobicity thus reducing affinity to the matrix. This feature needs to be balanced against the need for resolution though.

[0053] The use of the C-terminal sequencing agents to derivitise the free carboxyl groups which will reduce the adhesion between such peptides and a cation exchange resin. This may mean that cation exchange chromatography may be advantageous as a chromatographic separation step.

[0054] One can derivitise quite readily acetylate amine residues to achieve similar effects for anion exchange chromatography.

## Analysis of peptides by mass spectrometry:

### Ionisation Techniques:

[0055] In general peptide mixtures are injected into the mass spectrometer by electrospray or MALDI TOF, which leaves them in the vapour phase.

### Electrospray Ionisation:

[0056] Electrospray ionisation requires that the dilute solution of biomolecule be 'atomised' into the spectrometer from an insertion probe, i.e. in a fine spray. The solution is, for example, sprayed from the tip of a needle in an electrostatic field gradient. The mechanism of ionisation is not fully understood but is thought to work broadly as follows. The electrostatic field charges droplets formed at the probe tip promoting atomisation. In the stream of nitrogen the solvent is evaporated. With a small droplet, this results in concentration of the biomolecule. Given that most biomolecules have a net charge this increases the electrostatic repulsion of the dissolved protein. As evaporation continues this repulsion ultimately becomes greater than the surface tension of the droplet and the droplet 'explodes' into smaller droplets. The electrostatic field helps to further overcome the surface tension of the charged droplets. The evaporation continues from the smaller droplets which, in turn, explode iteratively until essentially the biomolecules are in the vapour phase, as is all the solvent.

### Atmospheric Pressure Chemical Ionisation:

[0057] An ionisation technique appropriate for use with LCMS, for analysing peptides is Atmospheric Pressure Chemical Ionisation (APCI). This is an electrospray based technique where the ionisation chamber is modified to include a discharge electrode which can be used to ionise the bath gas which in turn will collide with the vaporised sample molecules increasing ionisation of the

sample.

## Fast Atom Bombardment:

[0058]   This is an ionisation technique that is quite similar to APCI and is highly compatible with samples in solution. Typically, a continuous flow of liquid from a capillary electrophoresis column or an HPLC column can be pumped through an insertion probe to a hole or a frit at its tip where the solution is bombarded by accelerated atoms or ions, usually of xenon or caesium. Collision with the dissolved sample results in transfer of kinetic energy to and ionisation of the sample.

## Matrix Assisted Laser Desorption Ionisation (MALDI) :

[0059]   MALDI requires that the biomolecule solution be embedded in a large molar excess of an photo-excitable 'matrix'. The application of laser light of the appropriate frequency (266 nm beam for nicotinic acid ) results in the excitation of the matrix which in turn leads to excitation and ionisation of the embedded biomolecule. This technique imparts a significant quantity of translational energy to ions, but tends not to induce excessive fragmentation despite this. Accelerating voltages can again be used to control fragmentation with this technique though.

[0060]   MALDI techniques can be supported in two ways. One can embed proteins in a MALDI matrix, where the proteins themselves are not specifically excitable by laser or one can construct peptide labels that contain the necessary groups to allow laser energisation. The latter approach means the labels do not need to be embedded in a matrix before performing mass spectrometry. Such groups include nicotinic, sinapinic or cinnamic acid moieties. MALDI based cleavage of labels would probably be most effective with a photocleavable linker as this would avoid a cleavage step prior to performing MALDI mass spectrometry. The various excitable ionisation agents have different excitation frequencies so that a different frequency can be chosen to trigger ionisation from that used to cleave the photolysable linker. These excitable moieties are easily derivitised using standard synthetic techniques in organic chemistry so labels with multiple masses can be constructed in a combinatorial manner.

[0061]   All of the above techniques are routinely used with peptides and proteins and are preferred methods of ionisation with this invention.

## Mass Spectrometric Sensitivity and Quantitation of peptide tags:

[0062]   The end result is then that a population of proteins can be arbtitrarily sorted into populations of peptides of convenient size to be fed into a mass spectrometer for analysis. Completion of such an analysis for an entire cell's proteins would give a profile of what proteins

are present and in what relative quantities. Absolute quantitation could be acheived by 'spiking' a protein population with known quantities of particular proteins, known to be absent, e.g. plant proteins in animal samples or visa versa, against which to calibrate results. Internal quantities can be determined by measuring relative quantities of certain proteins present at relatively fixed concentrations in most cells such as histones. Various techniques coupled to certain mass spectrometer geometries permit good quantitation with a mass spectrometer. These issues are dealt with fully in GB 9719284.3.

## Mass Analyser Geometries:

[0063]   Mass spectrometry is a highly diverse discipline and numerous mass analyser configurations exist and which can often be combined in a variety of geometries to permit analysis of complex organic molecules such as the peptide tags generated with this invention.

## Accurate Mass Measurement:

[0064]   Double focussing mass spectrometers are capable of measuring molecular masses to a very high accuracy, i.e. fractions of a dalton. This permits one to distinguish molecules with identical integer mass but different atomic compositions with ease as fractional differences in the mass of different atomic isotopes allow such distinctions. For determining the molecular masses of a population of peptide tags, this technique may be very effective as it would allow identification of a significant proportion of peptides without requiring any sequencing even if some do have the same integral mass. The few ambiguous peptides that remain could be analysed by tandem mass spectrometry as discussed below.

## Sequencing of peptide tags by Tandem mass spectrometry:

[0065]   Peptides can be readily sequenced by tandem mass spectrometry. Tandem mass spectrometry describes a number of techniques in which a ions from a sample are selected by a first mass analyser on the basis of their mass charge ratio for further analysis by induced fragmentation of those selected ions. The fragmentation products are analysed by a second mass analyser. The first mass analyser in a tandem instrument acts as a filter selecting ions to enter the second mass analyser on the basis of their mass charge ratio, such that essentially a species of only a single mass/charge ratio, usually only a single peptide ion, enter the second mass analyser at a time. On leaving the first mass analyser, the selected peptide passes through a collision chamber, which results in fragmentation of the peptide. Since fragmentation occurs mostly at the peptide bond,

the pattern of fragments corresponds to a series of sub-species of peptides and amino acids that compose the original peptide. The distinct pattern of masses of single amino acids, 2-mers, 3-mers, etc. generated in the fragmentation of a peptide is sufficient to identify its sequence.

ION SOURCE -> MS1 -> COLLISION CELL -> MS2 -> ION DETECTOR

[0066] Various tandem geometries are possible. Conventional 'sector' instruments can be used where the electric sector provide the first mass analyser stage, the magnetic sector provides the second mass analyser, with a collision cell placed between the two sectors. This geometry is not ideal for peptide sequencing. Two complete sector mass analysers separated by a collision cell could be used for peptide sequencing. A more typical geometry used is a triple quadrupole where the first quadrupole filters ions for collision. The second quadrupole in a triple quadrupole acts as a collision chamber while the final quadrupole analyses the fragmentation products. This geometry is quite favorable. Another more favorable geometry is a Quadrupole/Orthogonal Time of Flight tandem instrument where the high scanning rate of a quadrupole is coupled to the greater sensitivity of a TOF mass analyser to identify the products of fragmentation.

### Sequencing with Ion Traps:

[0067] Ion Trap mass spectrometers are a relative of the quadrupole spectrometer. The ion trap generally has a 3 electrode construction - a cylindrical electrode with 'cap' electrodes at each end forming a cavity. A sinusoidal radio frequency potential is applied to the cylindrical electrode while the cap electrodes are biased with DC or AC potentials. Ions injected into the cavity are constrained to a stable circular trajectory by the oscillating electric field of the cylindrical electrode. However, for a given amplitude of the oscillating potential, certain ions will have an unstable trajectory and will be ejected from the trap. A sample of ions injected into the trap can be sequentially ejected from the trap according to their mass/charge ratio by altering the oscillating radio frequency potential. The ejected ions can then be detected allowing a mass spectrum to be produced.

[0068] Ion traps are generally operated with a small quantity of a 'bath gas', such as helium, present in the ion trap cavity. This increases both the resolution and the sensitivity of the device by collision with trapped ions. Collisions both increase ionisation when a sample is introduced into the trap and damp the amplitude and velocity of ion trajectories keeping them nearer the centre of the trap. This means that when the oscillating potential is changed, ions whose trajectories become unstable gain energy more rapidly, relative to the damped circulating ions and exit the trap in a tighter bunch giving a narrower larger peaks.

[0069] Ion traps can mimic tandem mass spectrometer geometries, in fact they can mimic multiple mass spectrometer geometries allowing complex analyses of trapped ions. A single mass species from a sample can be retained in a trap, i.e. all other species can be ejected and then the retained species can be carefully excited by super-imposing a second oscillating frequency on the first. The excited ions will then collide with the bath gas and will fragment if sufficiently excited. The fragments can then be analysed further. One can retain a fragment ion for further analysis by ejecting other ions and then exciting the fragment ion to fragment. This process can be repeated for as long as sufficient sample exists to permit further analysis. It should be noted that these instruments generally retain a high proportion of fragment ions after induced fragmentation. These instruments and FTICR mass spectrometers (discussed below) represent a form of temporally resolved tandem mass spectrometry rather than spatially resolved tandem mass spectrometry which is found in linear mass spectrometers.

[0070] For the purposes of protein profiling a peptide population, an ion trap is quite a good instrument. A sample of peptide tags can be injected into the spectrometer. Peptide tags that are expected to appear in a profile, such as housekeeping proteins or histone peptides from eukaryote cell samples, can be ejected specifically and quantified rapidly. The remaining peptides can be scanned. Totally new peptides can then be selectively retained from subsequent samples of the peptide population and can be induced to fragment allowing sequence data for that peptide to be acquired. Alternatively an Ion Trap can form the first stage of a tandem geometry instrument.

### Fourier Transform Ion Cyclotron Resonance Mass Spectrometry (FTICR MS):

[0071] FTICR mass spectrometry has similar features to ion traps in that a sample of ions is retained within a cavity but in FTICR MS the ions are trapped in a high vacuum chamber by crossed electric and magnetic fields. The electric field is generated by a pair of plate electrodes that form two sides of a box. The box is contained in the field of a superconducting magnet which in conjunction with the two plates, the trapping plates, constrain injected ions to a circular trajectory between the trapping plates, perpendicular to the applied magnetic field. The ions are excited to larger orbits by applying a radiofrequency pulse to two 'transmitter plates'which form two further opposing sides of the box. The cycloidal motion of the ions generate corresponding electric fields in the remaining two opposing sides of the box which comprise the 'receiver plates'. The excitation pulses excite ions to larger orbits which decay as the coherent motions of the ions is lost through collisions. The corresponding signals detected by the receiver plates are converted to a mass spectrum by fourier transform analysis.

**[0072]** For induced fragmentation experiments these instruments can perform in a similar manner to an ion trap - all ions except a single species of interest can be ejected from the trap. A collision gas can be introduced into the trap and fragmentation can be induced. The fragment ions can be subsequently analysed. Generally fragmentation products and bath gas combine to give poor resolution if analysed by FT of signals detected by the 'receiver plates', however the fragment ions can be ejected from the cavity and analysed in a tandem configuration with a quadrupole, for example.

**[0073]** For protein profiling FTICR MS could be used and may be advantageous as these instruments have a very high mass resolution allowing for accurate mass measurement so that peptides with the same integer mass but different atomic compositions can be resolved. Furthermore unidentified peptide tags can be subsequently analysed by fragmentation.

## Protein immobilisation:

**[0074]** A great deal of knowledge has been accumulated about specific protein chemistries particularly in the area of organic synthesis of peptides.

- R.B. Merrifield, Science 232: 341 -347, 1986.
- S.B.H. Kent, "Chemical Synthesis of Peptides and Proteins", Annu. Rev. Biochem. 1988. 57: 957 - 989.

## Linkers:

**[0075]** An important feature of this invention is cleavable linkers to their relevant biomolecules. Photocleavable linkers are particularly desirable as they allow for rapid, reagentless cleavage. For references, see:

- Theodora W. Greene, "Protective Groups in Organic Synthesis", 1981, Wiley-Interscience.

On photoremovable groups:

**[0076]**

- Patchornik, J. Am. Chem. Soc. 92: 6333 - , 1970.
- Amit et al, J. Org. Chem. 39: 192 - , 1974.

Liquid Chromatography:

**[0077]**

- R. Scopes, "Protein Purification: Principles and Practice", Springer-Verlag, 1982.
- M. Deutscher, "Guide to Protein Purification", Academic Press, 1990.

## Mass Spectrometry:

**[0078]** Electrospray mass spectrometry is the pre-ferred technique for sequencing peptides since it is a very soft technique and can be directly coupled to the liquid phase molecular biology used in this invention. For a full discussion of mass spectrometry techniques see:

- K. Biemann, "Mass Spectrometry of Peptides and Proteins", Annu. Rev. Biochem. 1992. **61**: 977 - 1010.
- R.A.W. Johnstone and M.E. Rose, "Mass Spectrometry for chemists and biochemists" 2nd edition, Cambridge University Press, 1996.

## Experiment

## Outline of embodiment of protein profiling

**[0079]** This comprises a system where

(i) a protein has its carboxyl groups protected, the last amino acid removed leaving just one carboxyl group free at the cleaved terminus.

(ii)This will be reacted with a biotinylation reagent, so that the carboxy terminus is labelled with biotin.

(iii)The protein is fragmented with a protease to leave peptide fragments, only the carboxyl one being biotinylated.

**[0080]** The biotin is used to attach the C terminal fragment to immobilised streptavidin, or preferably monomeric avidin, from which it can be released with mild acid and made available for MS - MS.

**[0081]** All reagents are available, and the chemistry is generally well-known as follows:

(i) The technique of carboxy-terminal sequencing of proteins is established. We note that the method of Boyd et al., (Boyd,, VL, Bozzini, M, Guga, PJ, DeFranco, RJ, Yuan, P-M, Loudon, GM and Nguyen, D; J. Org Chem, 60, 2581, (1995)) blocks the side chain carboxyls of aspartate and glutamate residues by amidation during removal of the terminal amino acid.

(ii) Biotinylation of the free carboxyl group at the carboxy terminus may be achieved using 5-(biotimamido)pentylamine/1-ethyl-3-[3-dimethylaminopropyl] carbodiimide hydrochloride, which is marketed by Pierce & Warriner (Lee, KY, Birckbichler, PJ and Patterson, MK, Clin Chem, 34, 906 (1988) for such a purpose.

(iii) Protease fragmentation of proteins on a membrane is an established technique (Sutton, CW, Pemberton, KS, Cottrell, JS, Corbett, JM, Wheeler, CH, Dunn, MJ and Pappin, DJ, Electrophoresis, 16,

308, (1995), and Millipore Corporation produce Im-mobilon®-CD and other PVDF membranes for that purpose. Monomeric avidin is produced by Pierce and Warriner, and allows release of biotinylated molecules using 2mM biotin in phosphate buffered saline.

[0082]   The remaining step in the method is the use of PVDF membranes (as used for trypsinisation) in lieu of Zitex® membranes for the sequencing reaction (I).

**Methodology**

**Binding of lysozyme to PVDF membrane**

[0083]   0.5mm squared pieces of PVDF (Millipore) were wetted with isopropanol and incubated in 20mg/ml lysozyme (Pharmacia) in PBS at room temperature for 30 minutes. The membranes were then air dried and stored at 4oC until used.

**Modification (carboxyl group protection) of lysozyme bound to PVDF**

[0084]   Modification solution was prepared by mixing 62mg of 2-ethyl-5-phenylisoxazolium-3'sulfonate (Aldrich) with 50ul of diisopropylethylamine (Aldrich) in 2mls of $CH_3CN$ 100ul of modification solution was added to each membrane and incubated at room tempera-ture for 4 hours.

[0085]   Following incubation 900ul of water was added and each membrane was gently shaken at room tem-perature for 30 minutes. Each membrane was then transferred to 50ul of $CH_3CN$, 450ul of water was added and the membranes were gently shaken at room tem-perature for 30 minutes.

[0086]   The each membrane was then transferred to 500ul of 2% trifluoroacetic acid and incubated at room temperature overnight

**Trypsin digest**

[0087]   Each membrane was transferred to 250ul of 25mM ammonium bicarbonate pH7.6 solution and gen-tly shaken at room temperature for 15 minutes.

[0088]   Each protein/protein containing membrane was added/transferred to 200ul of ammonium bicarbo-nate solution pH7.6 containing 5ug of trypsin and incu-bated at 37°C overnight.

**Eluation of protein/peptide fragments from membrane**

[0089]   Each membrane was transfered to 100ul of 50% formic acid/50% ethanol solution and incubated at room temperature for 30 minutes to remove the protein/peptides. The membranes were then removed and 300ul of water added to the 50% formic acid/50% etha-

nol solution containing the protein/peptides.

**Analysis**

[0090]   The following were analysed by reversed phase HPLC

[0091]   40ug of trypsin in PBS; 40ug of lysozyme in PBS; 40ug of lysozyme digested with trypsin; 40ug of trypsin digested with trypsin; membrane bound modified lysozyme digested with trypsin; membrane put through the modification protocol without lysozyme and digested with trypsin; membrane bound lysozyme unmodified and digested with trypsin; membrane bound lysozyme modified without trypsin digestion.

**Results**

[0092]   We have now performed the operation using PVDF membranes in lieu of Zitex® membranes for the sequencing reaction (I). We have found that the re-versed phase HPLC chromatogram for lysozyme (used as a typical protein) obtained after treatment with the sequencing reactions on a PVDF membrane and trypsinisation, from which the chromatogram for the same process in the absence of lysozyme has been subtracted, is similar to that obtained for lysozyme trypsinised directly. Hence the technologies are compat-ible and can be used to generate 'signature' peptides for MS-MS identification (data not shown).

**KEY TO THE DRAWINGS:**

**FIGURE 1**

[0093]

Step 1:   Extract proteins with harsh solvents and cap-ture extracted proteins onto a PVDF mem-brane

Step 2:   Loosely immobilised proteins can be washed to dispose of harsh solvents

Step 3:   Treat proteins with C-terminal coupling agent

Step 4:   Treat derivitised proteins with cyclisation re-agent and then cleave terminal amino acid from derivitised protein

Step 5:   Biotinylate newly exposed penultimate ami-no acid carboxyl group

Step 6:   Wash membrane bound proteins to remove chemical agents and cleave proteins with trypsin in physiological buffer

Step 7:   Capture terminal fragments onto avidinated beads

Step 8:   Wash away free peptides then release cap-tured peptide 'tags' for analysis

Step 9:   Analyse by MS or LC/MS/MS or MS/MS

**FIGURE 2**

[0094]

| | |
|---|---|
| Step 1: | Extract proteins with harsh solvents and capture extracted proteins onto a PVDF membrane |
| Step 2: | Loosely immobilised proteins can be washed to dispose of harsh solvents |
| Step 3: | Treat proteins with C-terminal coupling agent |
| Step 4: | Wash membrane bound proteins to remove chemical agents and cleave proteins with trypsin or other sequence specific cleavage agent in in physiological buffer |
| Step 5: | Biotinylate newly exposed carboxyl termini |
| Step 6: | Capture terminal fragments onto avidinated beads in an affinity column for example |
| Step 7: | Analyse eluted C-terminal by MS or LC/MS/MS or MS/MS |

**FIGURE 3**

[0095]

| | |
|---|---|
| Step 1: | Extract proteins with harsh solvents and capture extracted proteins onto a PVDF membrane |
| Step 2: | Loosely immobilised proteins can be washed to dispose of harsh solvents |
| Step 3: | Treat proteins with C-terminal coupling agent carrying immobilisation effector |
| Step 4: | Wash membrane bound proteins to remove chemical agents and cleave proteins with trypsin in physiological buffer |
| Step 5: | Capture terminal fragments onto avidinated beads |
| Step 6: | Wash away free peptides then release captured peptide 'tags' for analysis |
| Step 7: | Analyse by MS or LC/MS/MS or MS/MS |

**FIGURE 4**

[0096]

| | |
|---|---|
| Step 1: | Extract proteins with harsh solvents and capture extracted proteins onto a PVDF membrane |
| Step 2: | Loosely immobilised proteins can be washed to dispose of harsh solvents |
| Step 3: | Treat proteins with C-terminal coupling agent |
| Step 4: | Treat coupled proteins with derivitisation reagent to ensure all exposed carboxyls are capped |
| Step 5: | Treat derivitised proteins with cyclisation reagent and then cleave terminal amino acid from derivitised protein |
| Step 6: | Biotinylate newly exposed penultimate ami- |

| | |
|---|---|
| | no acid carboxyl group |
| Step 7: | Wash membrane bound proteins to remove chemical agents and cleave proteins with trypsin in physiological buffer |
| Step 8: | Capture terminal fragments onto avidinated beads |
| Step 9: | Wash away free peptides then release captured peptide 'tags' for analysis |
| Step 10: | Analyse by MS or LC/MS/MS or MS/MS |

**Claims**

1. A method for characterising polypeptides, which comprises :

(a) treating a sample comprising a population of a plurality of polypeptides with a cleavage agent which is known to recognise in polypeptide chains a specific amino acid residue or sequence and to cleave at a cleavage site, whereby the peculation is cleaved to generate peptide fragments;

(b) isolating a first population of peptide fragments which comprises only terminal peptide fragments bearing as a reference terminus the N-terminus or the C-terminus of the polypeptide from which they were fragmented, from a second population of peptide fragments which do not comprise the reference terminus, each peptide fragment of the first population bearing at the other end the cleavage site proximal to the reference terminus, said isolation step comprising immobilisation of the first population of peptide fragments or immobilisation of the second population of peptide fragments onto a solid support, and.

(c) determining by mass spectrometry a signature sequence of at least some of the isolated, fragments, which signature sequence is the sequence of a predetermined number of amino acid residues running from the cleavage site;

wherein a signature sequence characterise each polypeptide.

2. A method according to claim 1, wherein the reference terminus is attached to a solid phase support to immobilise the population of polypeptides or peptide fragments thereof.

3. A method according to claim 2, wherein the population of polypeptides is immobilised before treatment with the cleavage agent.

4. A method according to claim 2 or claim 3, wherein

the reference terminus is attached to the solid phase support by: (i) treating the polypeptides with a blocking agent to block all exposed reference groups, which comprise either carboxyl groups or primary amine groups; (ii) cleaving the reference terminal amino acids to expose unblocked reference termini; and iii) treating the unblocked reference termini with an immobilisation agent capable of coupling to the solid phase support; wherein step (b) comprises binding the treated reference termini to the solid phase support and removing unbound peptide fragments.

5. A method according to claim 1, which further comprises

(i) preparing the sample step (a) by pre-treating the polypeptides with a blocking agent to block all exposed reference groups, which comprise either carboxyl groups or primary amine groups, so that subsequent treatment of the sample with the cleavage agent generates peptide fragments bearing unblocked reference termini;
(ii) treating the unblocked reference termini with an immobilisation agent capable of coupling to a solid phase support ; and
(iii) binding the peptide fragments containing the unblocked reference termini to the solid phase support; wherein step (b) comprises eluting unbound peptide fragments therefrom.

6. A method according to claim 4 or claim 5, wherein the immobilisation agent comprises a biotinylation agent.

7. A method according to any one of claims 4 to 6, wherein the reference group is carboxyl.

8. A method according to any one of the preceding claims, wherein the cleavage agent comprises a peptidase.

9. A method according to any one of the preceding claims, wherein the sample of step (a) comprises a sub-cellular fraction.

10. A method according to any one of the preceding claims, which further comprises preparing the sample of step (a) by liquid chromatography.

11. A method according to any preceding claim, wherein the mass spectrometry is preceded by a high pressure liquid chromatography step to resolve the peptide fragments.

12. A method according to any one of claims 1 to 11, wherein the peptide fragments are subjected to ion

exchange chromatography before step (c).

13. A method according to any one of the preceding claims, wherein the predetermined number of amino acid residues is from 3 to 30.

14. A method for identifying polypeptides in a test sample, which comprises characterising the polypeptides in accordance with a method according to any one of the preceding claims, comparing the signature sequences and relative positions of the cleavage site obtained thereby with the signature sequences and relative positions of the cleavage site of reference polypeptides in order to identify the or each polypeptide in the test sample.

15. A method for assaying for one or more specific polypeptides in a test sample, which comprises performing a method according to any one of claims 1 to 14, wherein the cleavage agent and relative position of the cleavage site is predetermined and the signature sequence is determined in step (c) by assaying for a predetermined sequence of amino acid residues running from the cleavage site.

16. A method according to claim 15, wherein the cleavage site and signature sequence are predetermined by selecting corresponding sequences from one or more known target polypeptides.


**Patentansprüche**

1. Verfahren zur Charakterisierung von Polypeptiden, umfassend:

(a) Behandeln einer Probe, welche eine Population einer Vielzahl von Polypeptiden umfasst, mit einem Spaltungsmittel, welches bekanntermaßen in Polypeptidketten einen spezifischen Aminosäurerest oder eine Sequenz erkennt und an einer Schnittstelle schneidet, wobei die Population geschnitten wird, um Peptidfragmente zu erzeugen;
(b) Trennen einer ersten Population von Peptidfragmenten, welche nur terminale Peptidfragmente umfasst, die als Bezugsende den N-Terminus oder den C-Terminus des Polypeptids, von dem sie abgespalten wurden, tragen, von einer zweiten Population von Peptidfragmenten, welche nicht das Bezugsende umfassen, wobei jedes Peptidfragment der ersten Population am anderen Ende die Schnittstelle proximal zum Bezugsende trägt, wobei der Trennungsschritt die Immobilisierung der ersten Population von Peptidfragmente oder die Immobilisierung der zweiten Population von Peptidfragmente auf einem festen Träger um-

fasst, und

(c) Ermitteln einer Erkennugssequenz mittels Massenspektrometrie mindestens einiger isolierter Fragmente, wobei die Erkennungssequenz die Sequenz einer vorbestimmten Anzahl von Aminosäureresten ist, ausgehend von der Schnittstelle;

wobei eine Erkennungssequenz jedes Polypeptid charakterisiert.

2. Verfahren nach Anspruch 1, wobei das Bezugsende an einen Festphasenträger angebracht ist, um die Population von Polypeptiden oder Peptidfragmenten zu immobilisieren.

3. Verfahren nach Anspruch 2, wobei die Population von Polypeptiden vor der Behandlung mit dem Spaltungsmittel immobilisiert wird.

4. Verfahren nach Anspruch 2 oder Anspruch 3, wobei das Bezugsende an den Festphasenträger angebracht wird, mittels: (i) Behandlung der Polypeptide mit einem Blockierungsmittel, um alle exponierten Bezugsgruppen zu blockieren, welche entweder Carboxylgruppen oder primäre Aminogruppen umfassen; (ii) Schneiden der Bezugsenden-Aminosäuren, um unblockierte Bezugsenden zu exponieren; und (iii) Behandlung der unblockierten Bezugsenden mit einem Immobilisierungsmittel, geeignet zum Koppeln an den Festphasenträger; wobei Schritt (b) die Bindung der behandelten Bezugsenden an den Festphasenträger und die Entfernung ungebundener Peptidfragmente umfasst.

5. Verfahren nach Anspruch 1, welches weiterhin umfasst:

(i) Herstellen der Probe in Schritt (a) durch Vorbehandlung der Polypeptide mit einem Blockierungsmittel, um alle exponierten Bezugsgruppen zu blokkieren, welche entweder Carboxylgruppen oder primäre Aminogruppen umfassen, so dass eine darauffolgende Behandlung der Probe mit dem Spaltungsmittel Peptidfragmente erzeugt, welche unblockierte Bezugsenden tragen;
(ii) Behandlung der unblockierten Bezugsenden mit einem Immobilisierungsmittel, geeignet für das Koppeln an einen Festphasenträger; und
(iii) Binden der Peptidfragmente, welche die unblockierten Bezugsenden enthalten, an den Festphasenträger; wobei Schritt (b) die Elution von ungebundenen Peptidfragmenten umfasst.

6. Verfahren nach Anspruch 4 oder 5, wobei das Immobilisierungsmittel ein Biotinylierungsmittel umfasst.

7. Verfahren nach irgendeinem der Ansprüche 4 bis 6, wobei die Bezugsgruppe eine Carboxylgruppe ist.

8. Verfahren nach irgendeinem der vorangehenden Ansprüche, wobei das Spaltungsmittel eine Peptidase umfasst.

9. Verfahren nach irgendeinem der vorangehenden Ansprüche, wobei die Probe aus Schritt (a) eine subcelluläre Fraktion umfasst.

10. Verfahren nach irgendeinem der vorangehenden Ansprüche, welches weiterhin die Herstellung der Probe aus Schritt (a) mittels Flüssigkeitschromatographie umfasst.

11. Verfahren nach irgendeinem der vorangehenden Ansprüche, wobei der Massenspektrometrie eine Hochleistungsflüssigkeitschromatographie vorausgeht, um die Peptidfragmente aufzulösen.

12. Verfahren nach irgendeinem der Anprüche 1 bis 11, wobei die Peptidfragmente vor Schritt (c) einer Ionenaustauschchromatographie unterworfen werden.

13. Verfahren nach irgendeinem der vorangehenden Ansprüche, wobei die vorbestimmte Anzahl von Aminosäureresten von 3 bis 30 reicht.

14. Verfahren zur Identifizierung von Polypeptiden in einer Testprobe, umfassend die Charakterisierung der Polypeptide in Übereinstimmung mit einem Verfahren nach irgendeinem der vorangehenden Ansprüche, den Vergleich der Erkennungssequenzen und der relativen Positionen der dadurch erhaltenen Schnittstelle mit den Erkennungssequenzen und relativen Positionen der Schnittstelle von Bezugspolypeptiden, um die Polypeptide oder jedes Polypeptid in der Testprobe zu identifizieren.

15. Verfahren zur Untersuchung einer oder mehrerer spezifischer Polypeptide in einer Testprobe, umfassend die Durchführung eines Verfahrens nach irgendeinem der Ansprüche 1 bis 14, wobei das Spaltungsmittel und die relative Position der Schnittstelle vorbestimmt ist, und die Erkennungssequenz in Schritt (c) durch Untersuchung auf eine vorbestimmte Sequenz oder Aminosäurereste hin, ausgehend von der Schnittstelle, bestimmt wird.

16. Verfahren nach Anspruch 15, wobei die Schnittstelle und die Erkennungssequenz mittels Auswählen korrespondierender Sequenzen aus einem oder

mehreren bekannten Zielpolypeptiden vorbestimmt werden.

## Revendications

1. Procédé pour la caractérisation de polypeptides, comprenant :

> (a) le traitement d'un échantillon contenant une population formée d'une pluralité de polypeptides avec un agent de clivage dont on sait qu'il reconnaît dans les chaînes de polypeptides un résidu ou une séquence d'acide aminé spécifique et réalise le clivage sur un site de clivage, de façon à cliver la population pour produire des fragments de peptide ;
> (b) l'isolement d'une première population de fragments peptidiques composée uniquement de fragments peptidiques terminaux portant comme partie terminale de référence la partie N-terminale ou la partie C-terminale du polypeptide dont ils ont été extraits par fragmentation, par rapport à une seconde population de fragments peptidiques qui ne contiennent pas la partie terminale de référence, chaque fragment peptidique de la première population portant à l'autre extrémité le site de clivage proximal à la partie terminale de référence, ladite étape d'isolement comprenant l'immobilisation de la première population de fragments peptidiques ou l'immobilisation de la seconde population de fragments peptidiques sur un support solide, et
> (c) la détermination par spectrométrie de masse d'une séquence de signature de certains au moins des fragments isolés, laquelle séquence de signature est la séquence d'un nombre prédéterminé de résidus d'acide aminé à partir du site de clivage ;

> dans lequel une séquence de signature caractérise chaque polypeptide.

2. Procédé selon la revendication 1, **caractérisé en ce que** la partie terminale de référence est fixée à un support en phase solide afin d'immobiliser la population de polypeptides ou les fragments peptidiques de ceux-ci.

3. Procédé selon la revendication 2, **caractérisé en ce que** la population de polypeptides est immobilisée avant le traitement par l'agent de clivage.

4. Procédé selon la revendication 2 ou 3, **caractérisé en ce que** la partie terminale de référence est fixée au support en phase solide par :

> (i) traitement des polypeptides avec un agent de blocage pour bloquer tous les groupements de référence exposés, qui comprennent soit des groupements carboxyle, soit des groupements amine primaires ;
> (ii) clivage des acides aminés terminaux de référence de façon à exposer les parties terminales de référence non bloquées ; et
> (iii) traitement des parties terminales de référence non bloquées avec un agent d'immobilisation pouvant être couplé au support en phase solide ;

> dans lequel l'étape (b) comprend la liaison des parties terminales de référence traitées au support en phase solide et l'élimination des fragments peptidiques non liés.

5. Procédé selon la revendication 1, comprenant en outre :

> (i) la préparation de l'échantillon de l'étape (a) par prétraitement des polypeptides avec un agent de blocage de façon à bloquer tous les groupements de référence exposés, qui comprennent soit des groupements carboxyle, soit des groupements amine primaires, de telle sorte que le traitement subséquent de l'échantillon avec l'agent de clivage produise des fragments peptidiques portant des parties terminales de référence non bloquées ;
> (ii) le traitement des parties terminales de référence non bloquées avec un agent d'immobilisation pouvant être couplé à un support en phase solide ; et
> (iii) la liaison des fragments peptidiques contenant les parties terminales de référence non bloquées au support en phase solide ;

> **caractérisé en ce que** l'étape (b) comprend l'élution des fragments peptidiques non liés.

6. Procédé selon la revendication 4 ou la revendication 5, **caractérisé en ce que** l'agent d'immobilisation comprend un agent de biotinylation.

7. Procédé selon l'une quelconque des revendications 4 à 6, **caractérisé en ce que** le groupement de référence et un groupement carboxyle.

8. Procédé selon l'une quelconque des revendications précédentes, **caractérisé en ce que** l'agent de clivage contient une peptidase.

9. Procédé selon l'une quelconque des revendications précédentes, **caractérisé en ce que** l'échantillon de l'étape (a) contient une fraction subcellulaire.

**10.** Procédé selon l'une quelconque des revendications précédentes, **caractérisé en ce qu'il comprend en** outre la préparation de l'échantillon de l'étape (a) par chromatographie en phase liquide.

*5*

**11.** Procédé selon l'une quelconque des revendications précédentes, **caractérisé en ce que** la spectrométrie de masse est précédée d'une étape de chromatographie en phase liquide sous haute pression pour la résolution des fragments peptidiques.

*10*

**12.** Procédé selon l'une quelconque des revendications 1 à 11, **caractérisé en ce que** les fragments peptidiques sont soumis à une chromatographie par échange d'ions avant l'étape (c).

*15*

**13.** Procédé selon l'une quelconque des revendications précédentes, **caractérisé en ce que** le nombre prédéterminé de résidus d'acides aminés est compris entre 3 et 30.

*20*

**14.** Procédé pour l'identification de polypeptides dans un échantillon à tester, comprenant la caractérisation des polypeptides par un procédé selon l'une quelconque des revendications précédentes, la comparaison des séquences de signature et des positions relatives du site de clivage ainsi obtenues avec les séquences de signature et la position relative du site de clivage de polypeptides de référence afin d'identifier le polypeptide ou chaque polypeptide dans l'échantillon à tester.

*25*

*30*

**15.** Procédé pour la recherche d'un ou plusieurs polypeptides spécifiques dans un échantillon à tester, comprenant l'exécution d'un procédé selon l'une quelconque des revendications 1 à 14, **caractérisé en ce que** l'agent de clivage et la position relative du site de clivage sont prédéterminés et la séquence de signature est déterminée dans l'étape (c) par recherche d'une séquence prédéterminée de résidus d'acides aminés à partir du site de clivage.

*35*

*40*

**16.** Procédé selon la revendication 15, **caractérisé en ce que** le site de clivage et la séquence de signature sont prédéterminés par la sélection de séquences correspondantes parmi un ou plusieurs polypeptides cibles connus.

*45*

*50*

*55*
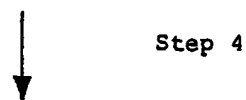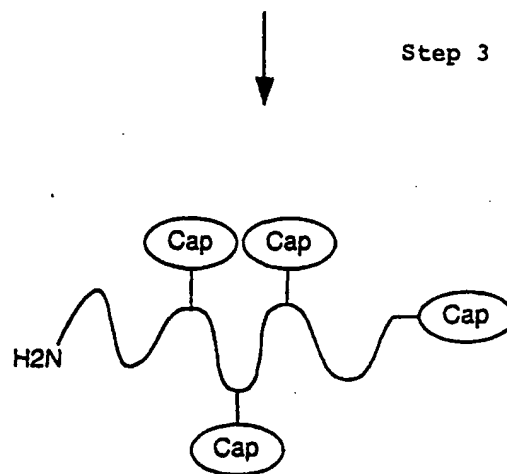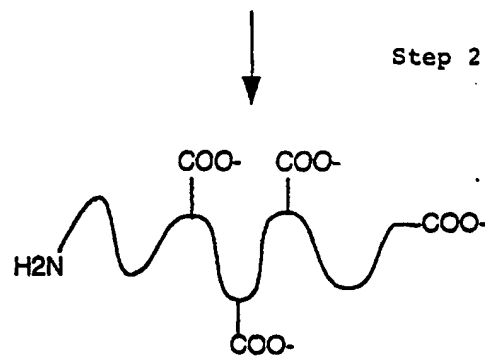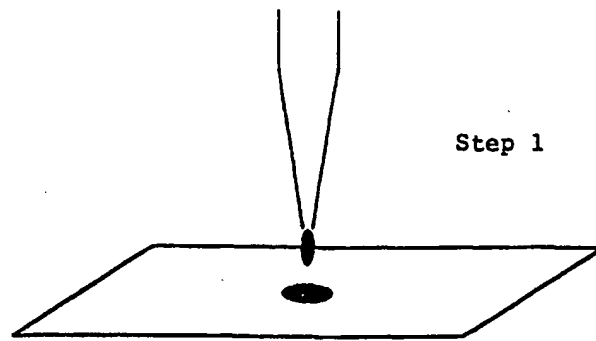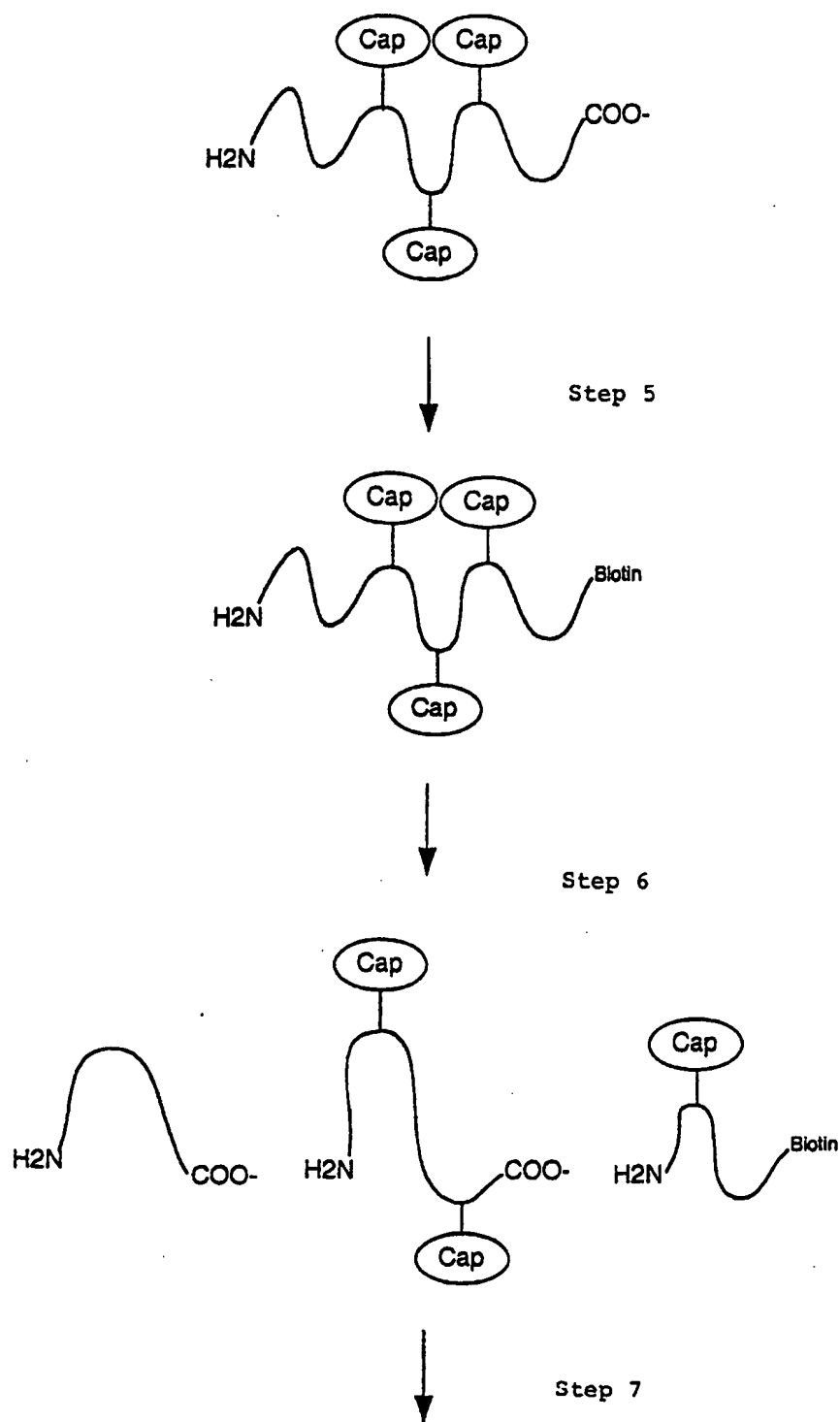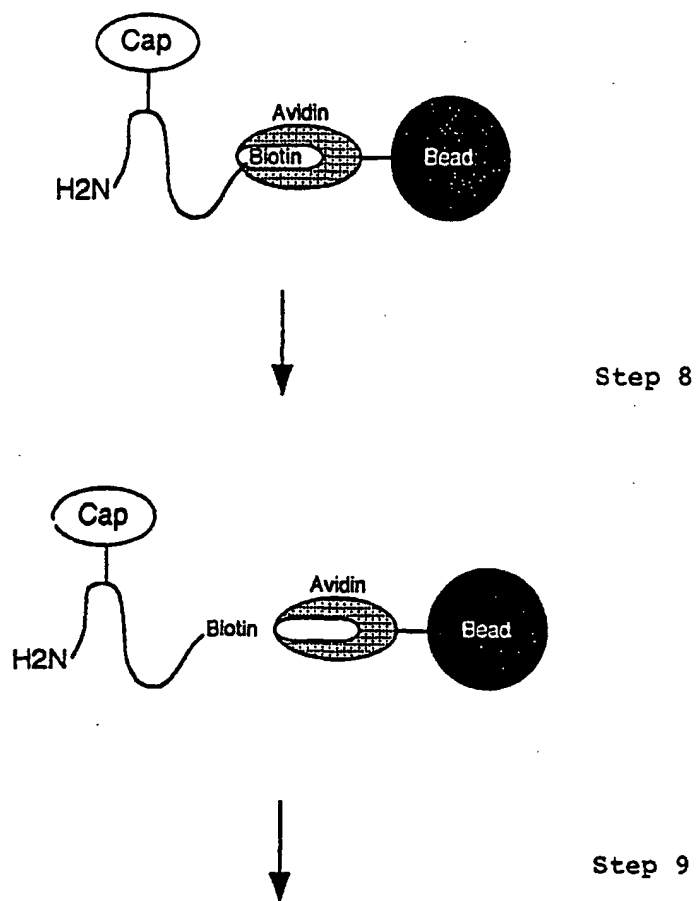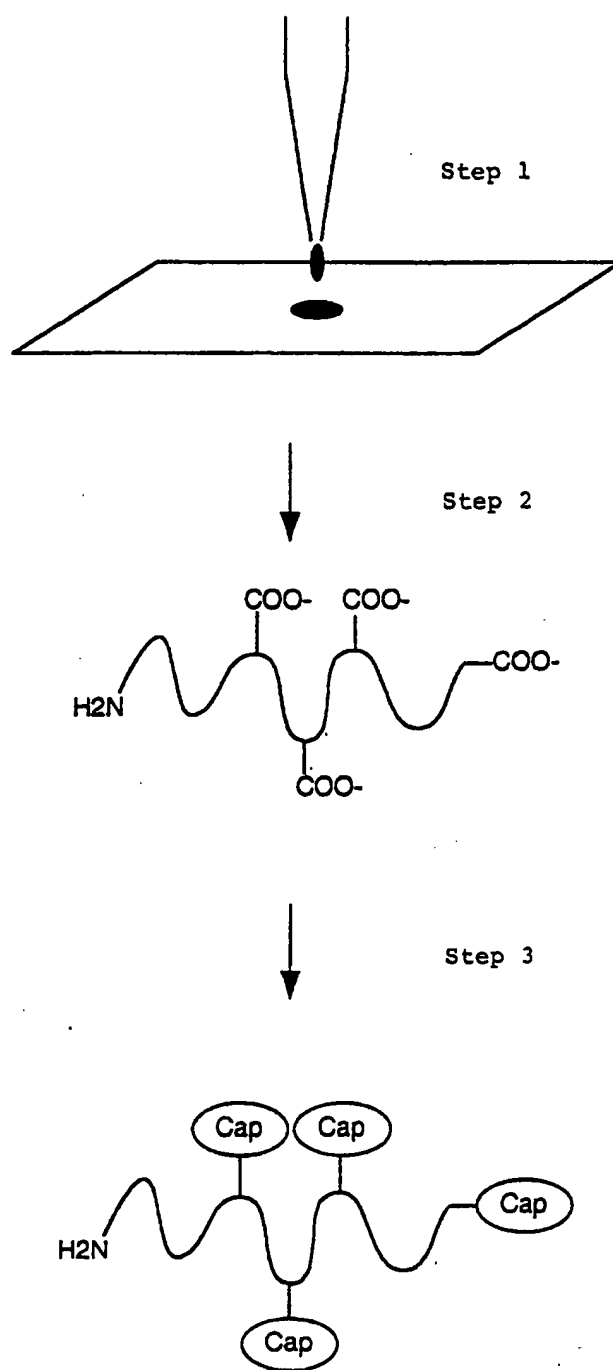
Step 1

Step 2

COO-  COO-

H2N

-COO-

COO-

Step 3

Cap   Cap

Cap

H2N

Cap

Step 4

FIGURE 1

Step 5

Step 6

Step 7

FIGURE 1 (continued)

Step 8



Step 9

FIGURE 1 (continued)

Step 1

Step 2

Step 3

FIGURE 2

Step 4

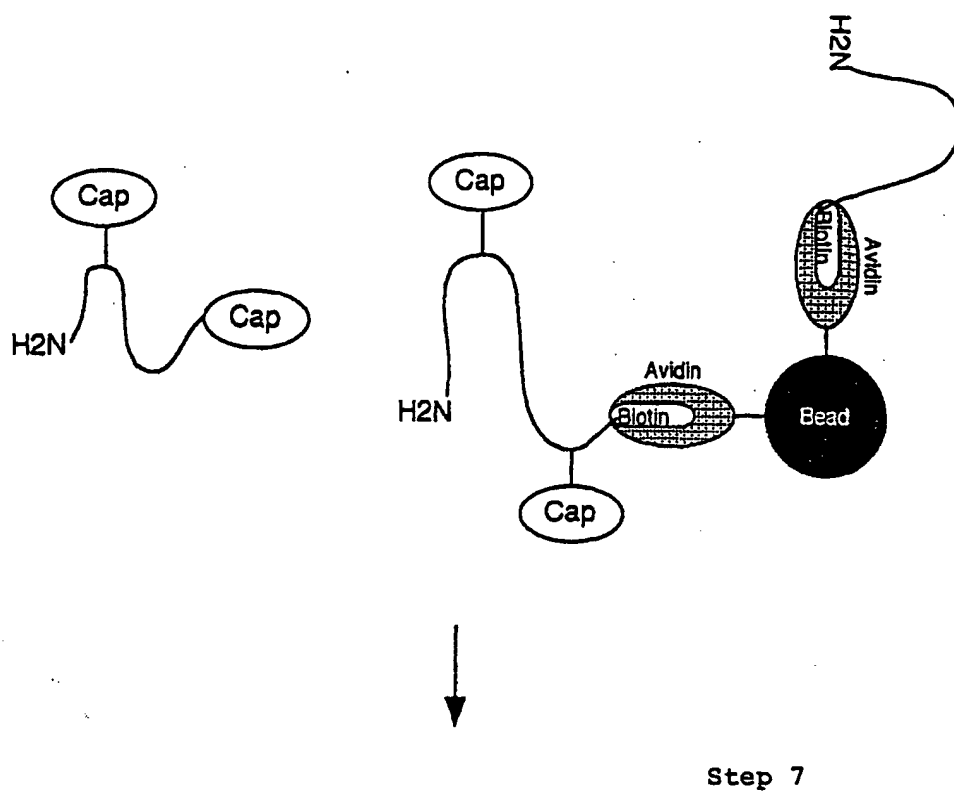Step 5

Step 6

FIGURE 2 (continued)

Step 7

FIGURE 2 (continued)

Step 1

Step 2

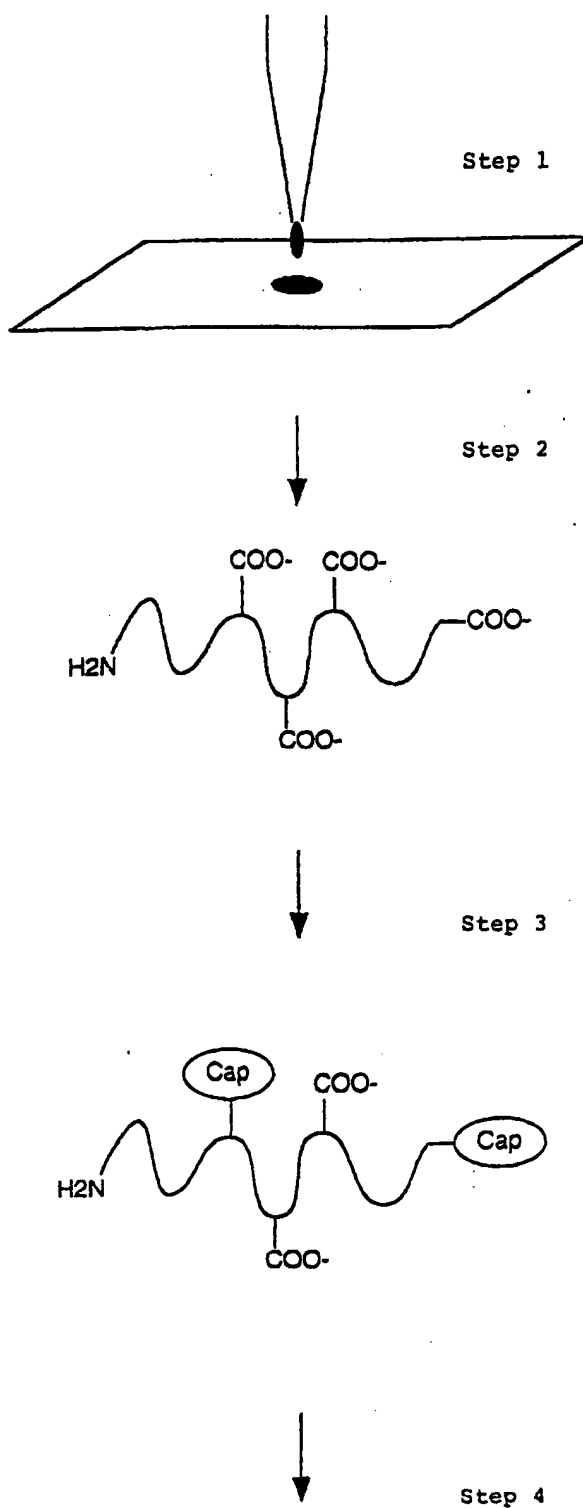COO-  COO-

H2N                    —COO-

COO-

Step 3

COO-  COO-

H2N          Cap   biotin

COO-

FIGURE 3

Step 4



Step 5



Step 6



Step 7

FIGURE 3 (continued)

Step 1

Step 2

COO-    COO-

                              -COO-

H2N

        COO-

Step 3

Cap    COO-

                    Cap

H2N

        COO-

Step 4
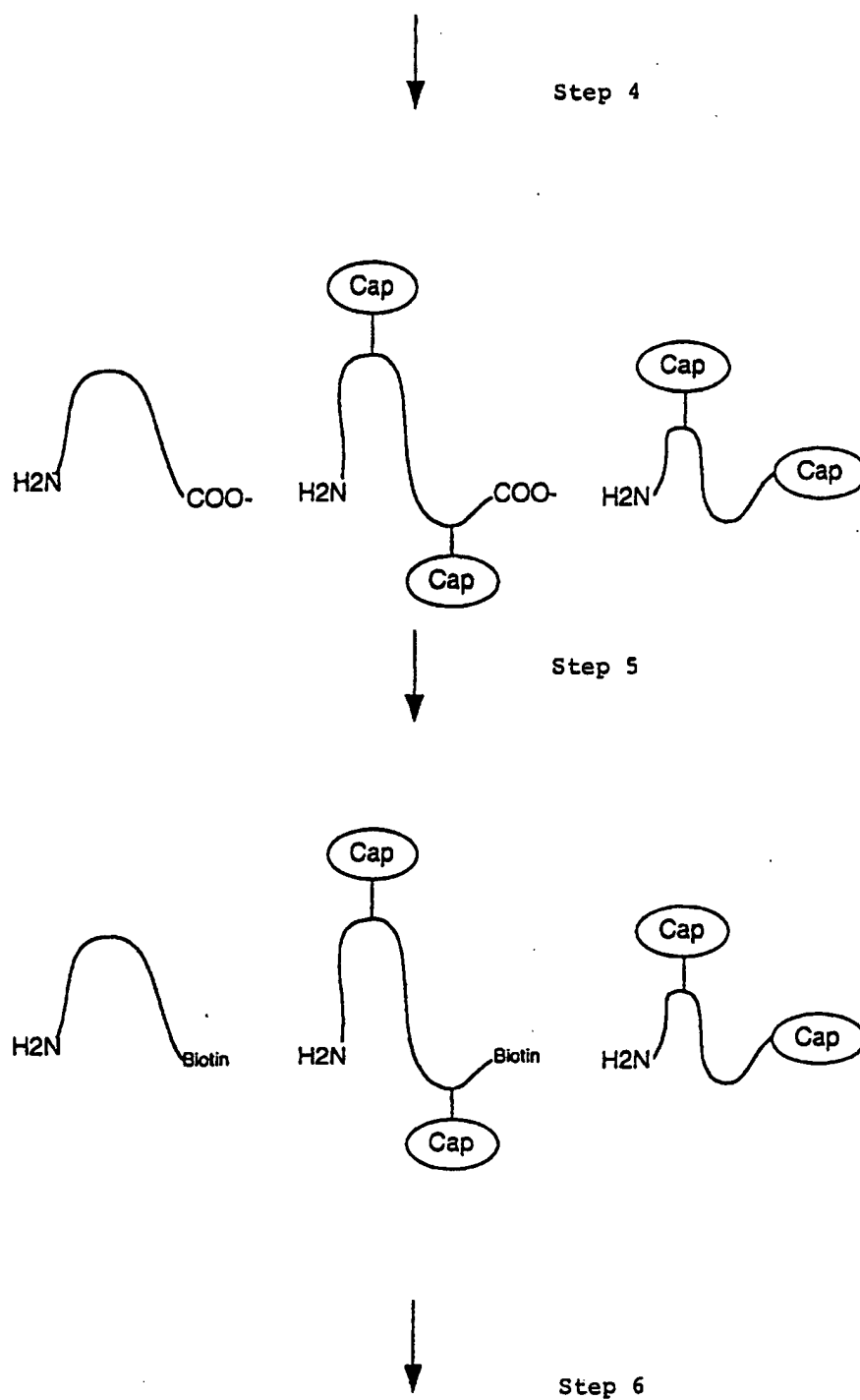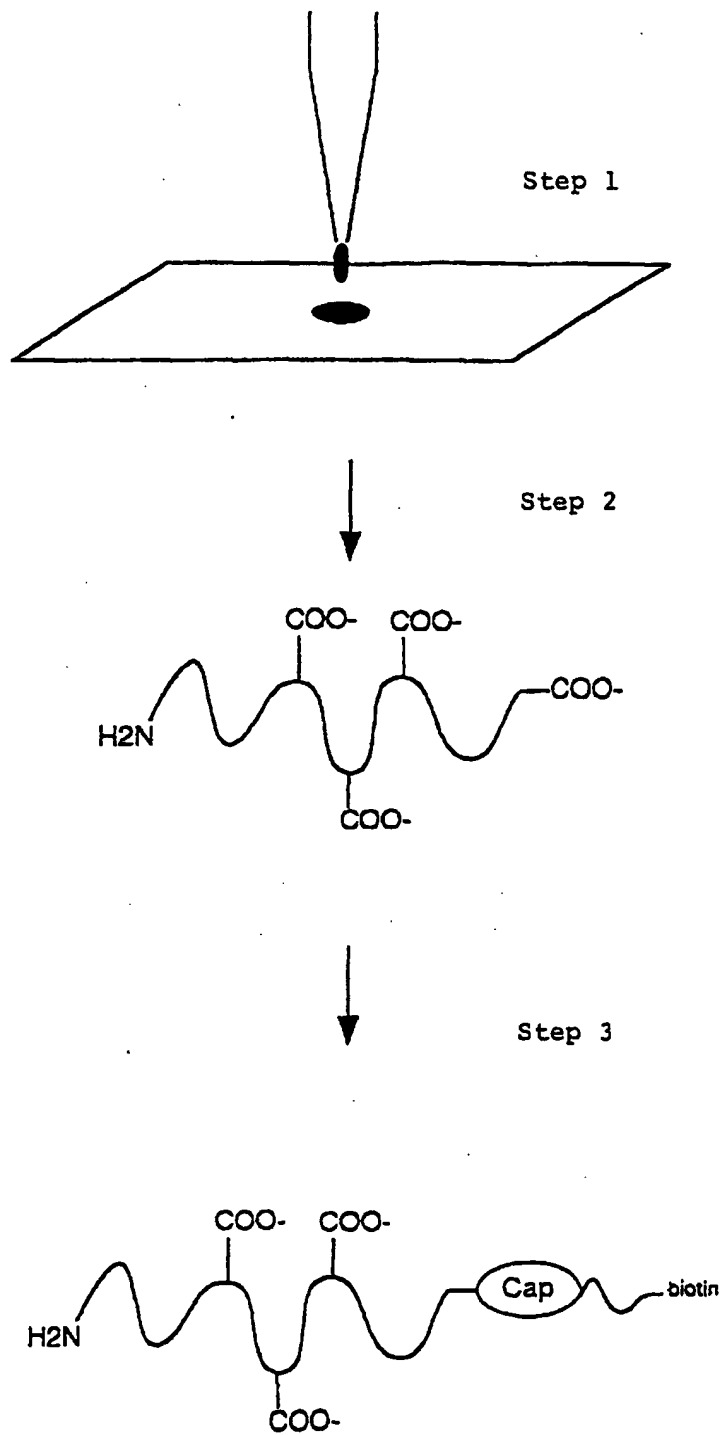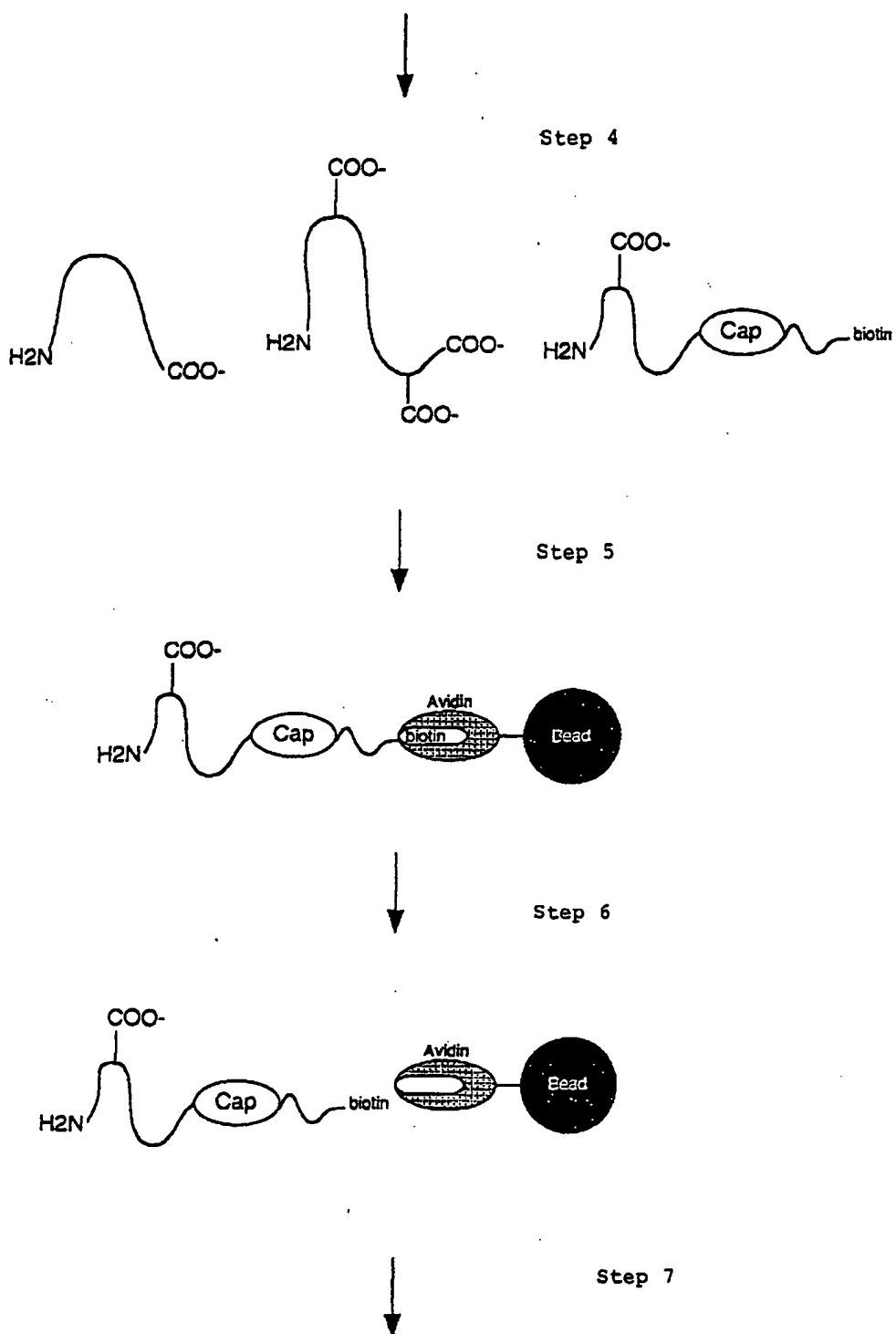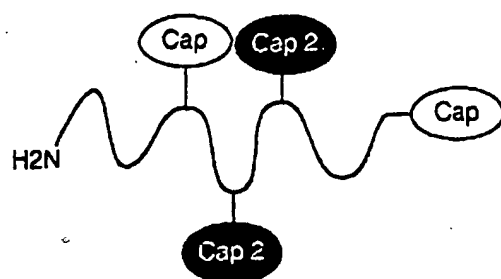
FIGURE 4

Step 5
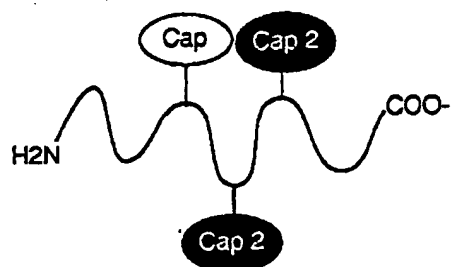


Step 6



Step 7

FIGURE 4(continued)

Step 8

Step 9

Step 10

FIGURE 4 (continued)